

Visual Analysis of the Paraiso Manifesto by Semantic Substrates

Aleks Aris*, Romain Vuillemot**

*University of Maryland, College Park

**LIRIS, INSA-Lyon, Université de Lyon, F-69621 Lyon, France

ABSTRACT

This paper provides one approach to analyze the social network of Paraiso Phenomenon by organizing nodes using their attributes (using NVSS [1]). Our conclusions include that the calls show a similar pattern to a normal distribution in terms of duration; that is, middle-length calls are frequent, while the shorter and longer calls are few. By assuming the importance of the longest calls, we spotted the locations of these calls in terms of originating cell tower. These indicate important places. Finally, call volume over ten days, day by day, indicate peak days on two of the four areas of the island.

KEYWORDS: Network visualization, semantic substrate design, information visualization, data exploration and analysis, graphical user interfaces.

INDEX TERMS: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Graphical User Interfaces (GUI); H.5.2 [Information Interfaces and Presentation]: User Interfaces—Screen design (e.g., text, graphics, color)

1 ANALYSIS

In analyzing the social network of the Paraiso phenomenon, we identified two ways of representing the data in terms of a network using NVSS ([1], [2]). These are:

- A) Nodes represent calls and towers. Links are between calls and towers and represent calls made using the connected tower.
- B) Nodes represent cell phones. Links are calls between cell phones.

There are trade-offs with each approach. (A) represents the data in its finest granularity (each call is visible as a node on the display). This allows to view which calls are made over the ten day period, day by day. However, it results in 10,000 nodes, which substantially slows NVSS. Also, nodes are harder to separate on the display and connections between cell phones are not available. A further disadvantage is that NVSS doesn't support multiple node types, i.e., nodes with different sets of attributes: calls and towers. However, combining the attributes of all node types and using this superset of attributes circumvents the limitation of NVSS and enables us to input the data to NVSS. On the other hand, (B) does not have this problem as the only node type is cell phone. Also, it results in only 400 nodes, which allows fast interactive exploration in NVSS. However, the granularity of calls is lost. In addition, NVSS permits only simple graphs; hence, links between two nodes have to be reduced to one link (by

aggregating the links between them). In addition, NVSS currently doesn't support link attributes; hence, this view would not be able to show any of the call attributes directly, such as duration, day of call, time of day, and tower. We could mitigate not being able to show the tower attribute by introducing towers as a second node type and connect the phone calls to the towers they used; however, even then we would not know the time attribute.

Considering these trade-offs, we chose to start using (A) and visualized the entire network in terms of calls and towers. Figure 1 shows this dataset. The top rectangular region has a map background and nodes on the map represent the towers. The bottom region contains the calls organized by day on the x-axis and duration (s) on the y-axis. A node represents a group of calls it represents. We see a normal distribution where the shorter and longer calls are few, while the middle-length calls are many. The longest calls (between 2101-2200 seconds) are selected using the duration filter on the Calls region and shows that these calls were made from the 1st and 11th tower on the 3rd day and from the 24th tower on the 9th day.

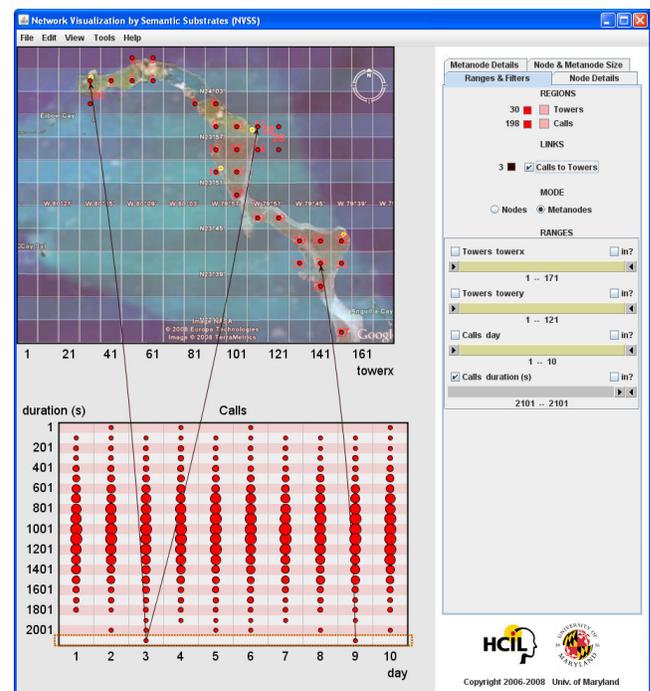


Figure 1. Longest calls occurred on the 3rd and 9th days.

Figure 2 shows the distribution of calls longer than 1900s in terms of the towers they used. There are 24 such calls over days 2-10.

The most used towers for these calls (longer than 1900s) are 30 and 1, 11 and 29, the left of the middle of the island (towers 9, 12, 13, 16) and the upper region of the lower part of the island (towers 20, 21, 22, and 24).

Aleks Aris: aris@cs.umd.edu

Romain Vuillemot: romain.vuillemot@insa-lyon.fr

LEAVE 0.5 INCH SPACE AT BOTTOM OF LEFT COLUMN ON FIRST PAGE FOR COPYRIGHT BLOCK

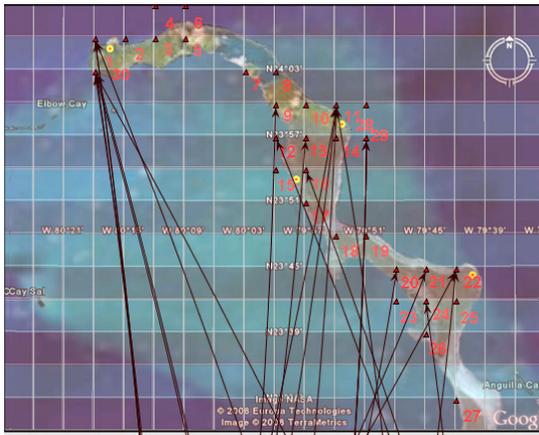


Figure 2. Calls longer than 1900s over 10 days.

Next, we analyzed the quantity of calls over the ten-day period in terms of 4 areas we defined on the map. These areas are spatially defined from top to bottom: top-left, top-middle, middle-bottom, and bottom-right. They contain the following towers:

- Top-Left: 1-6,30
- Top-Middle: 7-17, 28, 29
- Middle-Bottom: 18-26
- Bottom-Right: 27.

First, we focused on the Top-Left area and the calls made the 1st day from this area (Figure 3).

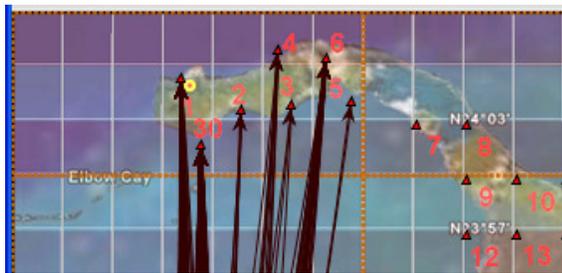


Figure 3. Focusing on the top-left region, analyzing the calls by tower on the 1st day

191 of the total 1765 calls made from this area were made in the 1st day. (The total number of calls from all areas is 9830.) The most active region in the area seems to be the leftmost, where towers 1 and 30 are located, and next is where the tower 6 is located. Also, no longest calls were made in the 1st day from this area.

Next, we browsed days 2-10 to see the number of calls on each day (Figure 4).

On the 2nd day, the left of the area (towers 1 & 30) increased in activity, while there is also increased activity at the top of the area (towers 4 & 6).

On the 3rd day, activity somewhat dropped in all areas and a bit increased on the 4th day.

On the 5th day, activity increased on the left of the area, namely towers 1 & 30, while the 6th day a peek is experienced on tower 1. On the following 2 days, it somewhat dropped and on the 10th day it increased a bit.

Next, we looked at the call volume in the middle-bottom area over 10 days (not illustrated) and inferred that tower 22 is heavily used for the first 5-6 days but less on the following 4-5 days.

We also looked at the bottom-right region for the call volume and detected no interesting patterns. It was active over all 10 days with no noticeable differences.

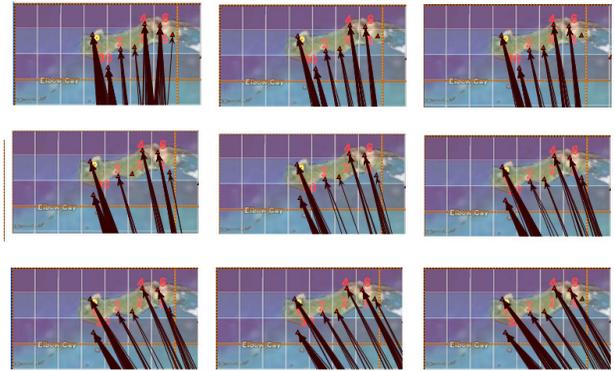


Figure 4. Focusing on the top-left region, analyzing the calls by tower on days 2-10 (top: 2,3,4, middle: 5,6,7, bottom: 8,9,10).

Next, due to the multitude of calls, we divided them to 5 categories in terms of their duration with boundaries of 400s, 800s, 1200s, 1600s, 2200s. Figure 5 shows the short calls (0-400s) on the left and the long calls (1600-2200s) on the right. While the top-middle region is active both in short and long calls, the top-left region is inactive in both calls except tower 30 and tower 1. Specifically, tower 30 is active in both, and tower 1 in long calls.

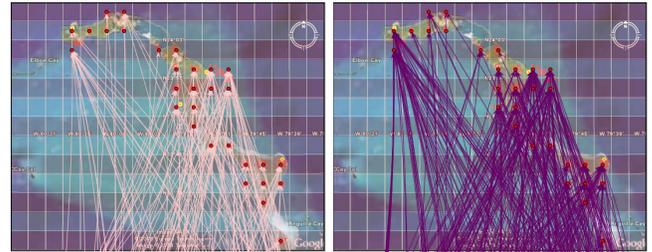


Figure 5. Contrasting short and long calls in terms of location of tower they use over the entire 10-day period.

Using approach (B), we looked at the social network of cell phones via NVSS and laid out nodes according to number of incoming and outgoing calls and formed 2 regions, one region to contain known id's (the id's whose name we know) and another region to contain unknown id's. As we discovered id's, we placed them into the known region. From Ferdinando Catalano (200), we found he talked to 6 people, one of which he talked 7 times (we assumed Esteban). From the number of incoming calls and connections to 2 other people that Ferdinando talked to, we found the Vidro family, David Vidro being the most active. We assumed the other two were Juan and Jorge. We labeled the other two ids (97 and 137) as friend of Ferdinando. Also, we noticed that id 0 receives the 2nd largest number of calls (210 calls (cell id=0); after David Vidro with 289 calls).

Acknowledgements

We appreciate the support of our faculty advisor, Ben Shneiderman.

REFERENCES

- [1] Ben Shneiderman, Aleks Aris, Network Visualization by Semantic Substrates, (*Proceedings of IEEE Visualization / Information Visualization*) *IEEE Trans. on Visualization and Computer Graphics* 12(5), 733-740, 2006.
- [2] Aris, A., Shneiderman, B., Designing Semantic Substrates for Visual Network Exploration, *Information Visualization Journal* 6(4), 281-300, 2007.